

# Convegno CCGRID 2002 a Berlino

**Claudio Arlandini, Maurizio Cremonesi**

*CILEA, Segrate*

## Abstract

Il secondo simposio internazionale della IEEE/ACM sul calcolo distribuito su cluster e via internet è stato un'interessante vetrina dello stato dell'arte del calcolo distribuito, che trova nello sviluppo di internet un'importante occasione di applicazione.

**Keywords:** Supercalcolo, Calcolo distribuito, Cluster computing, Grid computing.

Il convegno CCGRID [1] quest'anno si è svolto a Berlino dal 21 al 24 maggio, con la prima (mezza)giornata dedicata ai *tutorials*.

Il simposio si è svolto quindi su 3 giornate piene, organizzate in mezze giornate, ognuna introdotta da un *Keynote Address* di interesse generale, seguito da interventi più specialistici, divisi in 3 sessioni parallele. CCGRID 2002 è stato un convegno molto ricco, con interventi di alto livello espositivo, assicurato peraltro da una severa selezione preliminare degli articoli proposti.

CCGrid2002 è stato organizzato da ZIB Berlin, Freie Universität Berlin e la Humboldt-Universität zu Berlin. Il convegno si è svolto in un palazzo affacciato sulla Gendarmenmarkt, di fronte al Konzerthaus, tra il Französischer Dom e il Deutscher Dom, caratterizzati dalle alte torri rotonde.

Gli interventi presentati hanno riguardato gli aspetti più vari dell'elaborazione via rete, dai più modesti cluster di PC al calcolo cooperativo via internet tra centri di calcolo.

L'impressione globale è che ci sia un grande interesse sull'argomento e quindi una grande varietà di approcci e soluzioni originali a tutti i livelli. Inoltre l'interesse è tale da aver mosso alcune grandi aziende quali l'IBM, che è stata peraltro, nella persona di David Turek, degna ospite della cena, organizzata in un locale pittoresco di Berlino.

Guardando in retrospettiva l'evoluzione del calcolo intensivo, dai calcolatori vettoriali alle macchine massivamente parallele, sembra di

assistere ora ad un'analogica evoluzione. La tendenza odierna è di passare dal calcolo parallelo su cluster strettamente connessi ad una modalità di elaborazione basata su formidabili strumenti di calcolo che collaborano tra loro al fine di costituire una sinergia che permetta di sfruttare in modo complementare le particolari caratteristiche di ogni componente. Se un tempo la parallelizzazione del codice presentava formidabili sfide al programmatore, appena uscito dallo studio delle peculiarità del calcolo vettoriale, adesso il calcolo su rete territorialmente distribuita presenta difficoltà ma altrettanto grandi opportunità, per la potenza degli strumenti che possono essere sfruttati.

Tra gli interventi presentati è doveroso citare almeno i più significativi, partendo dalla presentazione che Charlie Katlett ha fatto del progetto TeraGrid [9], che coinvolge l'Argonne National Laboratory, il California Institute of Technology, il National Center for Supercomputing Applications e il San Diego Supercomputer Center ed è sostenuto dallo U.S. National Science Foundation. Il progetto, plurienale, dovrebbe portare alla realizzazione di un GRID, basato sulla cooperazione di cluster di workstations IA-64 con sistema operativo Linux realizzati mediante MPI e connessi utilizzando Globus Toolkit [10]. La potenza computazionale complessiva dovrebbe arrivare a quasi 14 TFlop/s con 750 TeraByte di dati online; la connessione tra i quattro centri do-

vrebbe arrivare ad avere una velocità di 40 Gbit/s.

Molti interventi hanno sollevato il problema della sicurezza dei collegamenti, evidentemente cruciale per l'elaborazione via internet. Il protocollo ritenuto più affidabile e più comodamente utilizzabile è SSH [8].

Non mancano soluzioni originali, quale quella presentata da Andrew A. Chien della Entropia Inc [2], che propone alle aziende l'utilizzo dei tempi morti dei PC dei dipendenti come importante risorsa di calcolo, basandosi sulle esperienze acquisite dalla compagnia in progetti scientifici, come la ricerca di nuovi numeri primi di Mersenne [5], che ha coinvolto migliaia di volontari in tutto il mondo. In particolare l'applicazione presentata si rivolgeva alle industrie farmaceutiche, per la ricerca di molecole attive durante lo sviluppo di nuovi farmaci. La sicurezza dei dati è comunque garantita, in quanto si inviano ai PC dati rigorosamente criptati e contenenti tracce che ne garantiscano la consistenza, al fine di evitare intercettazioni e manipolazioni dei risultati.

Diverse presentazioni hanno focalizzato l'attenzione al problema della gestione di file-system distribuiti più che sul calcolo. Uno dei progetti più interessanti al proposito è il Grid Datafarm [3], illustrato da Osamu Tatebe, dell'AIST, Giappone, che ha evidenziato i problemi posti e le soluzioni individuate per gestire enormi quantità di dati distribuiti geograficamente, a beneficio dei fisici delle alte energie.

Sempre restando nell'ambito della fisica, il progetto più ambizioso a livello mondiale, che richieda lo sviluppo di una rete complessa sia di calcolo che di gestione dati distribuiti è il progetto MONARC [11]. Come Hans F. Hoffman del CERN ha spiegato, si tratta in sostanza di realizzare un GRID a livello planetario, in grado di fornire potenza di calcolo e capacità di memorizzazione per riuscire ad analizzare i dati che saranno prodotti, a partire dal 2006, dagli esperimenti realizzati con l'LHC del CERN. Si stima che verranno prodotti dati al ritmo di alcuni PetaByte all'anno, che saranno interpretati e studiati dalla comunità mondiale dei fisici delle particelle.

Non mancano, come accennato sopra, le applicazioni più prettamente commerciali, come l'Océano autonomic system della IBM: Tamar

Eilam ha mostrato una demo del prototipo installato nel centro di ricerca IBM di New York. Il sistema si può descrivere come una Web server farm in grado di autoconfigurarsi dinamicamente secondo le necessità.

Sono stati presentati anche alcuni strumenti software per sviluppare programmi di calcolo parallelo su cluster. Uno di questi, P-GRADE, dell'MTA SZTAKI [6], Ungheria, per mezzo di un'interfaccia grafica, permette di realizzare programmi di calcolo distribuito su cluster funzionanti con PVM e MPI, ma senza doversi preoccupare dei dettagli inerenti l'uso di queste librerie di scambio messaggi.

Nel corso dell'ultimo giorno Daniel E. Atkins dell'"NSF Panel of Cyberinfrastructure" [7] e Kyriakos Baxevanidis della Commissione Europea [4], hanno presentato i programmi dei rispettivi enti per il sostegno ai progetti innovativi basati sul calcolo distribuito su reti regionali o continentali, a conferma dell'interesse verso l'argomento GRID, considerato strategico per la competitività delle aziende e delle istituzioni di ricerca.

### Bibliografia

- [1] CCGrid, URL: <http://www.ccgrid.org/>
- [2] ENTROPIA™ Distributed Computing, URL: <http://www.entropia.com/>
- [3] Grid Datafarm for Petascale Data Intensive Computing, URL: <http://datafarm.apgrid.org/>
- [4] IST Programme of DG INFSO, EC, URL: <http://www.cordis.lu/fp6/eoi-instruments/networks.htm>
- [5] MERSENNE Project, URL: <http://www.mersenne.org/>
- [6] MTA SZTAKI Laboratory of Parallel and Distributed Systems, URL: <http://www.lpds.sztaki.hu/>
- [7] NSF Directorate for Computer and Information Science and Engineering - Advisory Committee for Cyberinfrastructure, URL: [http://www.cise.nsf.gov/b\\_ribbon/index.html](http://www.cise.nsf.gov/b_ribbon/index.html)
- [8] SSH Communications Security, URL: <http://www.ssh.com/>
- [9] TeraGrid, URL: <http://www.teragrid.org/>
- [10] The globus project™, URL: <http://www.globus.org/>
- [11] The MONARC project, URL: <http://monarc.web.cern.ch/MONARC/>